IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Mechanism of Designing and Development of Speech Recognition Model.

Vidya Gangadhar Dahake and Dr. Jagdish D Kene

[1]Sri Satya Sai University of Technology & Medical Sciences, Sehore, India.

[2]Department of Electronics & Communication, RCOEM, Nagpur

Abstract

Study of speech recognition is a very challenging problem in its own right, with a well-defined set of applications. Speaker recognition is one of them an important and powerful application of speech recognition. The speaker recognition–verification and identification is a process of automatically recognizing who is speaking the basis of individual information included in speech signals. The process can be divided into speaker identification and speaker verification. Speaker identification determines which register speaker is providing a given utterance from amongst a set of known speakers. Speaker verification accept or reject the identity of a speaker, this process falls under biometric authentication.

Keywords: Speech detection, Microphone, Audio system, Software

Introduction

Speech recognition, or speech to text, involves capturing and digitizing the sound waves, converting them into basic language units or phonemes, constructing words from phonemes, and contextually analyzing the words to ensure correct spelling for words that sound like. The figure-.1 describes the on sight outline of the entire speech processing mechanism.[1]

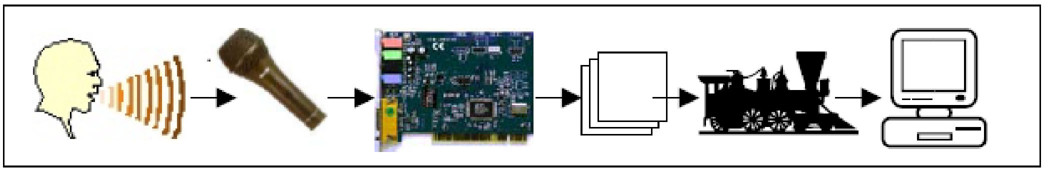


Figure-1 –Speech Processing Mechanism

Normally a speech recognition system consists of three subsystems (figure-2) i.e. that includes (1) microphone for translation of spoken words to analog signals, (2) An analog-to-digital signal processor and (3) software & hardware for translation of digital signal back to words.[2]

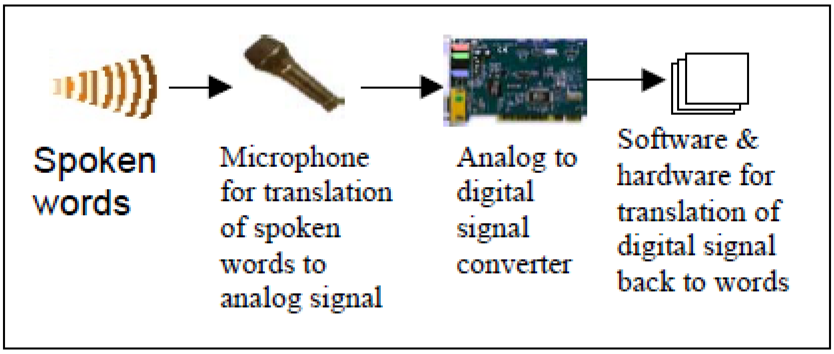


Figure-2 – Speech recognition system

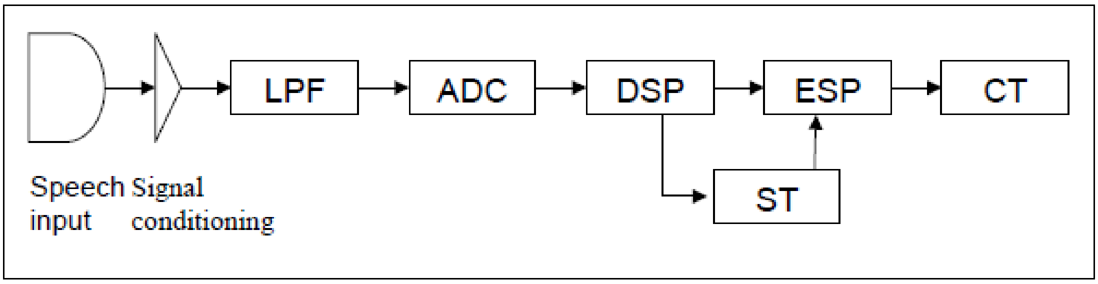
The process of conversion from speech to words is complex and varies slightly between systems. It consists of three steps (1) Feature extraction – Pre-processing of the speech signal, extracting the important features into feature vectors. (2) Phoneme recognition – bases on a statistically trained phoneme model (HMM) the most likely sequence of phoneme is calculated. (3) Word recognition – Based on statistically trained language model similar to the phoneme model, the most likely sequence of word are calculated. Based on this concept the researcher has designed and developed the speech recognition model to recognition the Indian Language i.e. alphabets of Gujarat Language. The proposed model is divided into two sub models (1) Preparation of speech file and (2) Speech dictation process[3]

Speech dictation process

After the preparation of master database of features of Gujarati Alphabets, the researcher has proposed the dictation model from where the actual human-machine interaction starts in the form of speech dictation. The researcher has divided the model into five different steps (figure –3) i.e. (1) Input acquition (2) Front end (3) Feature extractor (4) local match and (5) character printing. This division is, of course, somewhat arbitrary. In particular, the first two steps are frequently described as one system that produces features for classification stages. Here the functions are split out to emphasize their significance.[4]

Speech recognition system setup

Speech recognition setup is described as follow



The signal processing, begins with an input transducer, here it is a microphone. The signal conditioning circuit is to take the few mill volts of output from the input transducer and convert it to levels between 3 and 12 V. It also limits the input signal to prevent damage.[5]

References

1. Sadaoki, F 2005, '50 Years of Progress in Speech and Speaker Recognition Research', ECTI Transactions On Computer And Information Technology, vol.1, no. 2, pp. 64-74.
2. Anand, PS & Suneeta, A 2014, 'An Enhanced Cellular Automata based Scheme for Noise Filtering', International Journal of Signal Processing, Image Processing and Pattern Recognition vol. 7, pp. 231-242.
3. Cumani, S & Laface, P 2012, 'Analysis of Large-Scale SVM Training Algorithms for Language and Speaker Recognition', IEEE Transactions On Audio, Speech, And Language Processing, vol. 20, no. 5.
4. Hasan,T & Hansen, JHL 2014, 'Maximum Likelihood Acoustic Factor Analysis Models for Robust Speaker Verification in Noise', IEEE/ACM Transactions On Audio, Speech, And Language Processing, vol. 22, no.2, pp. 381-391.
5. Jiang, H & Deng, L 2001, 'A Bayesian Approach to the Verification Problem: Applications to Speaker Verification', IEEE Transactions On Speech And Audio Processing, vol. 9, no. 8.